



Universiteit
Leiden
The Netherlands

Machine learning and deep learning approaches for multivariate time series prediction and anomaly detection

Thill, M.

Citation

Thill, M. (2022, March 17). *Machine learning and deep learning approaches for multivariate time series prediction and anomaly detection*. Retrieved from <https://hdl.handle.net/1887/3279161>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3279161>

Note: To cite this publication please use the final published version (if applicable).



Universiteit
Leiden

Machine Learning and Deep Learning Approaches for Multivariate Time Series Prediction and Anomaly Detection

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 17 maart 2022
klokke 16.15 uur

door

MARKUS THILL

geboren te Patan, Nepal
in 1987

Promotiecommissie

Promotores: Prof.dr. T.H.W. Bäck

Prof.dr. W. Konen (Cologne University of Applied Sciences)

Co-promotor: Dr. H. Wang

Promotiecommissie: Prof.dr. A. Plaat

Prof.dr. K.J. Batenburg

Dr. M. Baratchi

Prof.dr. H. Trautmann (University of Münster)

Prof.dr. E. Alba (University of Malaga)

Abstract

In many real-world applications today, it is critical to continuously record and monitor certain machine or system health indicators in order to discover malfunctions or other abnormal behavior at an early stage and prevent potential harm. The demand for such reliable monitoring systems is expected to increase in the coming years. Particularly in the industrial context, in the course of ongoing digitization, it is becoming increasingly important to analyze growing volumes of data in an automated manner using state-of-the-art algorithms. In many practical applications, one has to deal with temporal data in the form of data streams or time series. The problem of detecting unusual (or anomalous) behavior in time series is commonly referred to as time series anomaly detection. Anomalies are events observed in the data that do not conform to the normal or expected behavior when viewed in their temporal context.

In the era of 'industry 4.0', 'cyber-physical systems' and 'big data', data-driven AI (artificial intelligence) approaches (designed to learn solely through data) from the fields of machine learning (ML), or deep learning (DL) – a sub-field of ML – have gained tremendous popularity. The majority of ML models are trained in a supervised fashion and require a labeled training data set. However, a particular problem in (time series) anomaly detection is that labeled data are usually relatively sparse and, as a consequence, supervised learning methods are mostly not feasible.

This thesis focuses on unsupervised machine learning algorithms for anomaly detection in time series. In an unsupervised learning setup, a model attempts to learn the normal behavior in a time series – which might already be contaminated with anomalies – without any external assistance. The model can then use its learned notion of normality to detect anomalous events. This work presents four unsupervised anomaly detection algorithms for multivariate time series, which can be used in different contexts: 1. SORAD learns an autoregressive prediction model and maintains a distribution of the prediction errors to detect anomalies. It can operate fully online and is up-and-running after a very short transient phase. Due to its online adaptability, it is advantageous for non-stationary time series. 2. DWT-MLEAD uses discrete wavelet transforms (DWT) to analyze a time series signal at different frequency scales to detect short-range and longer-range anomalies. It is also online-adaptable and computationally efficient. 3. LSTM-AD is a DL approach designed for longer time series, which uses a stack of recurrent long short-term memory (LSTM) neural networks to predict normal time-series behavior. Due to several extensions, it is particularly well-suited for quasi-periodic time series. 4. TCN-AE is a deep temporal autoencoder architecture, which – similarly to DWT-MLEAD – analyzes time series at different frequency scales and can learn short- and long-term relationships. It uses so-called dilated convolutional layers with learnable filters. TCN-AE is usually more compact in terms of model size than other DL models and can be trained faster due to its convolutional architecture.

We evaluate the algorithms presented in this work on challenging synthetic and real-world time series anomaly detection benchmarks, such as Mackey-Glass time series or electrocardiogram recordings, and compare them to other state-of-the-art algorithms.

Contents

Abstract	i
List of Tables	vi
List of Figures	vii
Chapter 1 Introduction	1
1.1 Background & Motivation	1
1.1.1 Introduction to Time Series Anomaly Detection	3
1.1.2 Challenges	4
1.1.3 Underlying Research Questions	6
1.2 Thesis Outline	8
Chapter 2 Fundamentals	9
2.1 Online Anomaly Detection	9
2.2 Scoring Process for Anomaly Detection Tasks	10
2.2.1 Anomaly Score & Anomaly Threshold	10
2.2.2 Anomaly Windows & Confusion Matrix	11
2.2.3 Algorithm Performance Measures	12
2.3 Time Series Anomaly Detection Benchmarks	13
2.3.1 Yahoo’s Webscope S5 Dataset	14
2.3.2 The Numenta Anomaly Benchmark	16
2.3.3 The MIT-BIH Arrhythmia Database	16
2.3.4 The Mackey-Glass Anomaly Benchmark	18
2.4 Time Series Characteristics	20
Chapter 3 Related Work	25
3.1 Algorithms used for Comparison Purposes	25
3.1.1 ADVec	25
3.1.2 NuPIC	26
3.1.3 LSTM-ED	26
3.1.4 DNN-AE	27
3.2 Other Anomaly Detection Algorithms	27
3.2.1 Online Algorithms	27
3.2.2 Deep Learning Approaches	28
3.3 Benchmarks	31
Chapter 4 SORAD: A Simple Online Regression Anomaly Detection Algorithm	33
4.1 Introduction	33



4.1.1	Related Work	34
4.2	Methods	35
4.2.1	Feature Generation using Sliding Windows	35
4.2.2	Offline Regression Anomaly Detection (Offline-RAD)	35
4.2.3	Online Estimation of a Distribution’s Mean and Variance	35
4.2.4	SORAD	37
4.3	Experimental Setup	38
4.3.1	Algorithm Setup	38
4.4	Results	40
4.5	Discussion	41
4.5.1	Transient Phase	41
4.5.2	Forgetting Factor	44
4.5.3	Detection Accuracy	46
4.5.4	Other algorithms	46
4.5.5	Limitations of SORAD	46
4.6	Conclusion	46
4.6.1	Possible Future Work	47
 Chapter 5 An Anomaly Detection Algorithm based on Discrete Wavelet Transforms		49
5.1	Introduction	49
5.1.1	Related Work	50
5.2	The Offline DWT-MLEAD Algorithm	51
5.2.1	Methods	51
5.2.2	Algorithms and their Settings	56
5.2.3	Results for the Offline DWT-MLEAD Algorithm	58
5.2.4	Discussion	60
5.2.5	Summary	60
5.3	Online-Adaptable DWT-MLEAD Algorithm	60
5.3.1	Online and Causal DWT & Sliding Windows	61
5.3.2	Online Estimation of the Mean and Covariance Matrix	61
5.3.3	Detecting Events in the DWT Tree and Anomaly Detection	62
5.3.4	Algorithmic Setup	62
5.3.5	Results	66
5.3.6	Discussion	67
5.4	Conclusion & Possible Future Work	69
 Chapter 6 Learning Quasi-periodic ECG Time Series with LSTM Networks		71
6.1	Introduction	71
6.1.1	Related Work	72
6.2	Methods	73

CONTENTS

6.2.1	LSTM for Time Series Prediction	73
6.2.2	Modeling the Residuals	75
6.2.3	Anomaly Detection	76
6.2.4	Window-Based Error Correction	77
6.3	Experimental Setup	78
6.3.1	The ECG-13 Benchmark	78
6.3.2	Parameterization of the Algorithms	78
6.4	Results & Analysis	79
6.5	Conclusion & Possible Future Work	83
Chapter 7 The Temporal Convolutional Autoencoder TCN-AE		85
7.1	Introduction	85
7.2	Methods	87
7.2.1	Intuition	88
7.2.2	Dilated Convolutions	89
7.2.3	Dilated Convolutional Layers in Neural Networks	92
7.2.4	Temporal Convolutional Networks	93
7.2.5	Unsupervised Anomaly Detection with TCN-AE	93
7.3	A Baseline Version of TCN-AE	94
7.3.1	The Baseline TCN-AE Architecture	94
7.3.2	Initial Experiments	95
7.3.3	Discussion	100
7.3.4	Summary	101
7.4	An Improved TCN-AE Architecture	101
7.4.1	Enhancements of the Baseline TCN-AE	101
7.4.2	Experimental Setup	105
7.4.3	Evaluation of TCN-AE on MGAB	107
7.4.4	Experiments, Results & Discussion for the ECG-25 Benchmark	107
7.5	Conclusion and Possible Future Work	121
Chapter 8 Conclusion and Outlook		123
8.1	Discussion	123
8.2	Conclusions	126
8.3	Future Work	128
Bibliography		129
Appendices		145
A	Extended Results for Chapter 7	147
B	Derivations	153



B.1	Batch Incremental Weighted Least Squares Estimator for multivariate Regression Tasks	153
B.2	Online Estimation of the Sample Mean and Covariance	157
B.2.1	The Weighted Mean and Covariance Matrix	157
B.2.2	Incremental Estimation with Exponentially Decaying Weights	162
B.2.3	The Covariance of Weighted Sample Means	166
B.2.4	Memory of the Exponentially Decaying Estimator	170
B.3	Relationship of the Mahalanobis Distance and the Chi-Square Distribution	173
B.3.1	Prerequisites	174
B.3.2	The Squared Mahalanobis Distance follows a Chi-Square Distribution	175
	Summary	179
	Samenvatting	181
	About the Author	183

List of Tables

2.1	Confusion Matrix	12
2.2	Anomaly types in the ECG-13 and ECG-25 data	17
4.1	Results for various algorithms on the Yahoo S5 datasets	44
4.2	Results for online algorithms after transient phase	45
4.3	Computation times of the algorithms on datasets A1–A4	45
5.1	Results for various algorithms on the A3 and NAB dataset	57
5.2	Computation times of the algorithms on datasets A3 and NAB	58
5.3	Results for various algorithms on the datasets A1–A4 and NAB	67
6.1	Summary of the the parameters used for the LSTM anomaly detector	79
6.2	Results for the ECG-13 dataset	80
6.3	Results with and without the window-based-error correction	82
6.4	Various metrics for 5 different anomaly classes	82
6.5	Results when LSTM is trained only with nominal sequences	83
7.1	Results for MGAB	100
7.2	Common parameters of the NN-based anomaly detection algorithms	106
7.3	Results for MGAB	107
7.4	Summary of all TCN-AE variants	109
7.5	Impact of the individual TCN-AE components for the ECG-25 data	111
7.6	Summary of the results for the ECG-25 data	114
7.7	Number of detected anomalies for the algorithms, by anomaly type	116
7.8	Computation times	116
7.9	F_1 -scores of the algorithms on all 25 time series of the ECG-25 benchmark	117
8.1	Suitability of some anomaly detection algorithms for various characteristics	125
A.1	Summary for the ECG-25 data when tuning threshold on 10% of the data	147
A.2	F_1 -scores of the algorithms, when tuning threshold on 10% of the data	148
A.3	TCN-AE components’ impact when tuning the threshold on 10% of the data	148
A.4	F_1 -scores of the TCN-AE variants on all 25 ECG-25 time series	151

List of Figures

1.1	An example time series from an industrial machine	2
2.1	A time series taken from the Numenta Anomaly Benchmark	11
2.2	Example plot for the A1 data	15
2.3	Example plot for the A2 data	15
2.4	Example plot for the A3 data	15
2.5	Example plot for the A4 data	15
2.6	Example ECG signal from the MIT-BIH Arrhythmia database	18
2.7	A section of a Mackey-Glass time series containing three anomalies	19
2.8	Time delay embedding of the Mackey-Glass attractor	20
2.9	Example for the creation of a Mackey-Glass time series with a temporal anomaly	22
2.10	Mackey-Glass time series, revealing the anomalies in Fig. 2.7	23
4.1	The ϵ -quantiles $\pm z_\epsilon$ of $\mathcal{N}(0, s_0^2)$	37
4.2	Time series with the anomalies detected by different algorithms	41
4.3	Multiobjective plot for different algorithms	42
4.4	Multiobjective plot for different SORAD variants	42
4.5	The performance of SORAD (without forgetting) over a range of thresholds	43
4.6	Comparing results on A4 for various forgetting factors of SORAD	43
5.1	Example of a decimating DWT using Haar Wavelets	52
5.2	Detecting anomalies with leaf counters	54
5.3	Example time series taken from the NAB data algorithm's detections	59
5.4	Multiobjective plot for the NAB and A3 dataset	59
5.5	Detecting anomalies with causal leaf counters	64
5.6	Example time series	65
5.7	Multiobjective plot for Yahoo's Webscope S5 benchmark and the NAB	68
6.1	LSTM-AD architecture	74
6.2	Gaussian fit without and with the removal of outliers	76
6.3	Window-based error correction method	77
6.4	Training vs. test errors (MSE) for the individual time series	79
6.5	Precision-Recall plot for LSTM-AD, NuPIC and ADVec	81
6.6	Example ECG time series with anomaly detections of the algorithms	81

LIST OF FIGURES

7.1	Frequency responses of several filters taken from 4 layers of TCN-AE	91
7.2	Architecture of the baseline TCN-AE	96
7.3	Encoding of MG time series by TCN-AE and t-SNE	98
7.4	Encoding MG time series into a 3d-vector	99
7.5	The architecture of TCN-AE's encoder	103
7.6	The architecture of TCN-AE's decoder	104
7.7	Reconstruction of MLII and V1 of ECG signal #1	108
7.8	Activations inside the trained final TCN-AE model	110
7.9	Heatmap, comparing TCN-AE to other algorithms	111
7.10	Precision-recall curves for the algorithms on the ECG-25 data	115
7.11	Erroneous reconstruction leading to a false-positive	119
7.12	Two segments taken from ECG time series #33	120
A.1	Tuning variants of TCN-AE on different subsets of the ECG-25 data	149
A.2	Tuning all algorithms on different subsets of the ECG-25 data	149
A.3	Histogram of the reconstruction error	150
B.1	Example: Comparison of the distributions two sample means	171